Roteiro da Prática - Aula 21 de Setembro de 2012

## PARTE I

1. Acesse a base Iris a partir da pagina da UCI (*http://archive.ics.uci.edu/ml/datasets/Iris*)
   1. Quantos atributos têm a base?
   2. Existe algum atributo com valor ausente?
   3. Qual o número de classes?
   4. Você conseguiu entender o que cada atributo representa?
2. Clique em *Data Folder*. Clique com o botão direito em *iris.data* e salve o arquivo.
3. Colocar o arquivo no formato *.arff* (Weka na Munheca)
4. Aplicar *3-nn (lazy.iBk)* e *percentage split* com 70%**.**
5. Anotar a acurácia.

6. Remover os atributos referentes à pétala
7. Aplicar 3-*nn*
8. Anotar a acurácia.

9. Remover os atributos referentes à sépala
10. Aplicar 3-*nn*
11. Anotar a acurácia.

12. O que aconteceu com a acurácia quando os atributos pétala e sépala foram removidos? Você acha que eles são relevantes para a classificação?

13. Remover os atributos referentes a *width*
14. Aplicar *3-nn*
15. Anotar a acurácia.

16. Remover os atributos referentes a *length*
17. Aplicar 3-*nn*
18. Anotar a acurácia.

19. O que aconteceu com a acurácia quando os atributos referentes a largura e altura foram removidos? Você acha que ambos são relevantes para a classificação? Remova atributo por atributo e identifique o que você considera menos relevante.

## PARTE II

20. Aplicar o algoritmo de agrupamento *k*-means (weka.clusteres.simpleKMeans)
    1. Utilizar os valores para *k* iguais a {2,3,4,5}.
    2. Selecionar *Classes to clusters evolution (Nom - class)*
    3. Anotar a acurácia

    4. Qual o valor de *k* faz mais sentido utilizar? Você acha que a acurácia é uma boa medida para validar os grupos obtidos?